Naval Health Research Center

RUNNING PERFORMANCE AS AN INDICATOR OF VO_{2max}: DISTANCE EFFECTS

Ross R. Vickers, Jr.

Report No. 01-20

NAVAL HEALTH RESEARCH CENTER
P.O. BOX 85122
SAN DIEGO, CALIFORNIA 92138

NAVAL MEDICAL RESEARCH AND DEVELOPMENT COMMAND
BETHESDA, MARYLAND

Running Performance as an Indicator of VO_{2max}: Distance Effects

Ross R. Vickers, Jr.

Human Performance Program
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92186-5122

e-mail: Vickers@nhrc.navy.mil

Report Number 01-20, supported by the Office of Naval Research, Arlington, VA, under Work Unit No. 63706N M0096.001-6417 and by the U. S. Army Medical Materiel and Research Command under Work Unit No. Army-Reim-60109. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government.

This research has been conducted in compliance with all applicable Federal Regulations governing the protection of human subjects. No human subjects were directly involved in this review.

Approved for public release, distribution unlimited.

EXECUTIVE SUMMARY

Background

Running performance often is used to evaluate aerobic capacity. Run tests are a useful alternative to laboratory measures of oxygen uptake because running performance is related to those measures. Run tests provide less precise estimates of aerobic capacity than laboratory measurement, but are much easier to conduct. The use of run tests, therefore, is a tradeoff between precision of estimation and simplicity of administration. Run tests must meet some minimum standard of estimation precision to justify their use.

Objective

This report reviews the literature relating aerobic capacity to running performance. The goal was to construct a model to predict run test estimation precision based on the distance or duration of the test. The model could answer questions such as "How long must a run be to provide a valid indication of aerobic capacity?" and "How much will precision increase if a 3-mile run is used instead of a 1.5-mile run?"

Approach

The published literature was searched to identify studies of maximal oxygen uptake (VO_{2max}) and running performance. A meta-analysis was conducted on reported correlations between VO_{2max} and performance extracted from 122 studies.

Results

The correlation between VO_{2max} and performance increased with distance, but only up to a point. For fixed-distance runs, the size of the correlation increased up to 2 km, then remained constant. For fixed-time runs, the correlation appeared to be constant for runs of 12 min or longer. Above these cutoffs, the fixed-time correlation (r = .797) was slightly higher than the fixed-distance (r = .718) correlation. These figures indicate a standard error between ~3.7 and ~4 $ml \bullet kg^{-1} \bullet min^{-1}$ compared to a range of ~2.5 to ~3 $ml \bullet kg^{-1} \bullet min^{-1}$ for laboratory tests.

Conclusions

Run tests should be at least 2 km in distance or 12 min in duration to maximize validity as indicators of aerobic capacity. Increasing distance or time beyond these minimum values does not improve run test validity as an indicator of VO_{2max}. Other things equal, fixed-time tests are preferable to fixed-distance tests. These tests estimate aerobic capacity with reasonable precision.

Introduction

Maximum oxygen uptake (VO_{2max}) is an important indicator of cardiorespiratory function (McArdle, Katch, & Katch, 1996). Laboratory tests that measure oxygen uptake during treadmill runs or cycle ergometry are the gold standard for assessing this capacity. These tests require special equipment and significant time investments to assess a single subject. The resource intensive character of the tests makes simpler alternatives attractive in many situations.

Run tests are a popular alternative means of estimating VO_{2max}. For example, run tests are common in fitness assessments of school children and military personnel. There is strong empirical justification for the use of run tests as a less technology intensive, cost-effective substitute for laboratory measures of VO_{2max}. Prior reviews of VO_{2max} and running performance (Baumgartner & Jackson, 1982; Knapik, 1989; Safrit, Hooper, Ehlert, Costa, & Patterson, 1988) identify numerous studies that reported VO_{2max}-running performance correlations that typically are between r = .50 and r = .80. Safrit et al. (1988) computed an average of r = .741, after correcting for measurement error, for runs covering 1 mile or more or lasting 9 minutes or longer.[1]

Previous reviews clearly demonstrate that run tests can be valid indicators of aerobic capacity, but an important question has not been examined in detail: "Do longer runs provide better estimates of aerobic capacity?" If so, the longest distance that the test population can complete should be used to estimate aerobic capacity. Also, some populations (e.g., children, elderly) may not be able to complete a long enough run to obtain acceptable VO_{2max} estimates. If so, some other method must be used to assess aerobic capacity. An examination of the relationship between run distance and validity will answer important questions such as "What is the shortest run that will meet a specified validity criterion (e.g., r = .65)?" and "How much would be gained if the current test were lengthened by 500 meters?" The answers to these questions have important implications for the effective use of run tests, particularly in applied settings.

Run test validity as an indicator of aerobic capacity[2] should increase with distance. Aerobic processes provide an increasing proportion of the total energy for performance as run distance increases (Spencer & Gastin, 2001. Increased dependence on aerobic energy should make the rate at which aerobic energy can be generated increasingly important for performance. Individual differences in that rate (i.e., differences in aerobic capacity) should be more strongly related to performance for longer runs. It follows that VO_{2max}, an indicator of aerobic capacity, should be more strongly related to performance for longer runs.

Two lines of evidence support the above arguments. First, studies that assessed performance for several distances have found stronger correlations for longer runs (Burke, 1976; Farrell, Wilmore, Coyle, Billing, & Costill, 1979; Shaver, 1975; Weyand, Cureton, Conley, Sloniger, & Liu, 1994). Second, mathematical models based on world records predict that a 1% difference in aerobic capacity yields a 0.3% performance difference at 400 m, but a 0.997% difference at 10 km (Ward-Smith, 1999). The close correspondence between aerobic capacity differences and performance differences at longer distances should translate into a stronger association between the two at longer distances.

The only quantitative review of the VO_{2max}-running performance literature contradicted the prediction that validity increases with run distance. Safrit et al. (1988) found no difference between shorter and longer runs in their analysis. However, their review only included runs ≥ 1 mile or ≥ 9 minutes. Most of the increase in the proportion of energy derived from aerobic processes occurs for shorter runs (Ward-Smith, 1999). The proportion increases from ~7% at 100 m to ~74% at 1500 m, then to ~97% at 10 km. If validity parallels dependence on aerobic energy, about 75% of the expected increase in validity coefficients was not covered in Safrit et al.'s (1988) review.

This review extends the quantitative analysis of run test validity initiated by Safrit et al. (1988). Meta-analysis (Hedges & Olkin, 1985; Hunter & Schmidt, 1990) is used to evaluate quantitative models relating distance/duration to run test validity. The review covers runs from 10-m sprints to 84.4-km ultramarathons.

Methods

Literature Search

The literature search was conducted in a series of steps designed to ensure broad coverage of published and unpublished research:

1. Articles cited by Safrit et al. (1988), Baumgartner and Jackson (1982), and Knapik (1989) formed the initial list of studies.
2. The Medline, PsychLit, and Discus databases were searched to identify additional studies using "maximal oxygen uptake" with "run time" or "running" as the primary keywords. Additional searches were performed with "maximum oxygen uptake," "maximal oxygen capacity," "aerobic capacity," and "VO_{2max}" as variations on maximal oxygen uptake. "Performance" was used as an alternative to run time.

3. The articles identified in steps 1 and 2 were examined. Those articles that reported at least 1 relevant correlation were retained.
4. An ancestry search (Rosenthal, 1984; White, 1994) was conducted by examining the reference lists in the articles retained in step 3.
5. Year-by-year searches were conducted in *Journal of Sports Medicine and Physical Fitness*, *Medicine and Science in Sports and Exercise*, *European Journal of Applied Physiology*, and *Research Quarterly for Sports and Exercise*. Each journal contributed multiple articles in steps 1 through 4. All volumes of the first two journals were reviewed; the latter two journals were reviewed from 1975 to present.
6. The Naval Health Research Center and San Diego State University library catalogues were searched to identify unpublished studies (e.g., Master's theses, Doctoral dissertations).
7. The PubMed database was searched to identify any additional publications appearing during the time that references were being collected. The Arelated articles≅ option of the program was examined for each new article found. This step updated the ancestry search.

The search produced 130 relevant studies, but only 122 were used in the analyses. Six studies (Butts, Henry, & McLean, 1991; Kohrt, Morgan, Bates & Skinner, 1987; Kohrt, O'Connor & Skinner, 1988; Krahenbuhl, Wells, Brown, & Ward, 1979; Schabort, Killian, St Clair Gibson, Hawley, & Noakes, 2000; Zhou, Robson, King, & Davis, 1997) were dropped because the run was one of several physically demanding activities performed in sequence. Fatigue from the other activities might affect the validity coefficients. The study by Cureton, Sloniger, O'Bannon, Black, and McCormack (1995) was dropped because it pooled data from several investigations. Other reports based on parts of the data included more detail on procedures and participant characteristics. The additional detail was useful for analyzing sources of variation in run test validity. The study by Doolittle and Bigbee (1968) was dropped because it reported a rank-order correlation rather than a Pearson product-moment correlation.

The remaining 122 studies reported results for 156 distinct samples. Because participants in some studies ran more than one distance, a total of 273 correlations were available based on VO_{2max} data from 6,140 individuals paired with 10,173 run performances.

Data Extraction

The information extracted from each report consisted of the sample size, the type of run test (fixed-distance or fixed-time), the distance run, the average run time, and the VO_{2max}-running

performance correlation. Performance was recorded a number of different ways in different studies. Performance on fixed-distance tests was usually recorded as a run time, but sometimes was represented by average running velocity. Performance on fixed-time tests typically was recorded as distance, but sometimes was reported as a predicted VO_{2max}. VO_{2max} predictions usually were computed using equations that involved only run distance. However, in some cases the predictions were based on multivariate equations with other predictors such as weight or gender.

Two steps were taken to make sure that correlations were comparable across studies. All of the correlations that used run time as the performance criterion were reversed. For each of the other criteria, higher values indicated better performance. The correlations, therefore, were nearly all positive. When run time was the criterion, lower scores indicated better performance and nearly all correlations were negative. Reversing the sign for these correlations meant that the results from all studies were expressed using coefficients that indicated how strongly VO_{2max} was related to good performance.

The second step taken to ensure that correlations were comparable restricted the set of results for the estimated VO_{2max} criterion. Correlations between measured and estimated VO_{2max} were included in the review only if the prediction equation was a linear function of distance with no other predictors. When these conditions are met, the prediction is merely a linear transformation of distance. Linear transformations of variables produce correlations that are identical to those for the variable itself (Hays, 1963), so the correlation between VO_{2max} and predicted VO_{2max} would be identical to the correlation between VO_{2max} and distance. This identity did not apply in studies where other predictors (e.g., weight, gender) or higher powers of distance (e.g., distance squared) were included in the predictive equation.

A separate record was constructed for each run test in a study. Thus, if a study included 1500-m, 5-km, and 10-km runs, a separate record was constructed for each distance. Sample attributes were duplicated on each record. Each record was treated as a separate case in the analysis. This decision meant that the cases analyzed were not entirely independent, thereby introducing statistical complexities for significance testing (Becker & Schram, 1994). The common meta-analytic practice of averaging effect sizes to produce a single value for each sample was not suitable for the present purposes. Averaging would have prevented meaningful analysis of the relationship between validity and test length.

Data Analysis

Rosenthal and DiMatteo (2001) capture the intended spirit of the present data analysis with two observations: "Meta-analysis is not inherently different from primary data analysis; it requires the same basic tools, thought processes, and cautions" (Rosenthal & DiMatteo, 2001, p. 78). "The best quality scientific exploration is often one that poses unadorned, straightforward questions and uses simple statistical techniques for analysis" (Rosenthal & DiMatteo, 2001, p. 68). A meta-analysis can appear complex because it involves a number of decision points (Wanous, Sullivan, & Malinak, 1989) and because effect sizes are analyzed rather than raw data. However, the essential computational procedures are analogous to familiar procedures for computing descriptive statistics, analysis of variance (ANOVA), and regression. The central components of the procedures in this paper were:

A. Olkin and Pratt's (1958) correction for sample bias in the estimated correlations was applied. Hedges and Olkin (1985) note that this correction is most important when 0.4 # r # 0.6 and sample size is small (e.g., n < 15). The average correlation reported by Safrit et al. (1988) was just above the upper end of this range, and many of the correlations reviewed (66 of 273, 24.2%) were from samples with n # 15. These figures suggested that the unbiased correlations should be used to protect against underestimating the true population correlation.
B. Fisher's *r*-to-*z* transformation was applied to normalize the distribution of correlations. The data points analyzed and predicted, therefore, are labeled $z_{UF(i)}$ as a reminder that they are **u**nbiased, **F**isher-transformed estimates of the population correlations for a given sample, denoted by the "*i*" in the subscript.
C. Each reported correlation was compared to a predicted value (i.e., $z_{UF(i)} - z_{UF}'$). The predicted values were familiar elements of standard analysis procedures. For example, the predicted values in one analysis of variance model were the means for all tests of specific distances (e.g., 800 m, 1500 m). The predicted values in another analysis were determined from the regression of $z_{UF(i)}$ on the logarithm of distance.
D. The difference between the observed and predicted values was standardized. This was accomplished by dividing $z_{UF(i)} - z_{UF}'$ by the standard deviation for the transformed correlation (i.e., $1//(N_i - 3)$).
E. The standardized value for the difference was squared to produce a Π^2 with 1 degree of freedom (Hays, 1963).
F. The Π^2 values for all correlations in the analysis were summed to produce an overall Π^2 that was the summary fit statistic for the model.

G. The Π^2 values for competing models were compared to determine which model best accounted for the observed variation in the correlations.

This summary shows that the computations involve differences between observed and predicted values. The differences are directly comparable to the deviations and/or residuals computed for descriptive statistics, ANOVA, or regression analyses of raw data. The statistical comparisons between models are comparable to using incremental variance explained to select a model in primary data analyses.

Meta-analysts must choose between fixed-effects and random-effects models (Hedges & Olkin, 1985; Hedges & Vevea, 1998; Raudenbush, 1994). Fixed-effects models were the starting point for the analyses, but a random-effects model was the end point. Fixed-effects models have smaller error variances than random-effects models (Becker & Schram, 1994; Erez, Bloom & Wells, 1996; Hedges & Vevea, 1998). Smaller error variance means larger standardized differences for fixed-effects analyses than for random-effects analyses. The overall model Π^2 is the sum of the squared standardized values (Hays, 1963), so underestimating error variance increases Π^2. This fact makes fixed-effects analyses lenient relative to random-effects models. However, fixed-effects models are a necessary first step in the iterative computation of the random-effects variance estimate in any case. Hedges and Vevea's (1998) procedures were used to compute a random-effects model after using fixed-effects analyses to choose between models. This decision made it possible to compare the models directly because each model was being used to account for the same Π^2. Hedges and Vevea's (1998) Equation 10 was used to compute the random-effects component of variance following the initial fixed-effects analysis.

Analyses were conducted with the general linear model (GLM) and linear regression procedures in SPSS-PC (SPSS, Inc., 1998a,b). The weighted least squares option in each procedure was used to apply the (n - 3) weight. Using this weighting option, the sums of squares reported in the analysis results are Π^2 values equal to Hedges' Q (cf., Hedges & Olkin, 1985, pp. 235-241). The GLM procedure was used for analyses involving discrete groups (e.g., males and females) and for multivariate models. Linear regression was used for analyses of nominally continuous variables (e.g., age). Nominally continuous variables were covariates in the multivariate models.

Model Comparison and Selection

Statistical significance tests are an imperfect guide to model selection (Morrison & Henkel, 1970; Harlow, Mulaik, & Steiger, 1997). Even very small effects are statistically

significant when examined in large samples (Rosenthal & Rosnow, 1984). Including weak effects in a model increases parametric complexity with little gain in predictive accuracy. Thus, the question of whether the increase in explanatory power justifies the increased complexity of the model. Identification of a parsimonious model, therefore, involves a tradeoff between explanatory power and complexity (Popper, 1959; Mulaik et al, 1989).

Two steps were taken to foster parsimony. Hoelter's (1983) critical N, the smallest sample size for which an observed difference would be statistically significant, was applied. If critical N is large, the effect arguably is too small to be important. Hoelter's (1983) rule of thumb that critical N should be > 200 was adopted to identify effect sizes too small to be practically or theoretically important.

The second protection against unnecessarily complex models was based on goodness of fit statistics for the model (cf., Arbuckle & Wothke, 1999; Bentler & Bonnet, 1980, Bollen, 1989, for discussions of goodness of fit). The Tucker-Lewis index (TLI, Tucker & Lewis, 1973) was adopted as a goodness-of-fit indicator:

$$TLI = (\Pi_N^2/df_N - \Pi_M^2/df_M)/(\Pi_N^2/df_N - 1)$$

where *N* indicates the null model and *M* indicates the alternative model. The expected value of Π^2 is 1.00 when chance is the only source of variation, so TLI was the proportion of the greater than chance variation in the observed correlations accounted for by a model. James, Mulaik, and Brett's (1982) parsimony adjustment then was applied:

$$PTLI = TLI*(df_M/df_N)$$

Basically, PTLI increases when the proportional gain in explanatory power exceeds the proportional decrease in degrees of freedom. Mulaik et al. (1989) explain the rationale for this adjustment in detail.

Results

Fixed-distance and fixed-time tests were considered separately. This approach avoided confounding cases in which distance or time was an experimental design variable defining the run test with cases where the same variables were performance indices.

Fixed-distance Tests

General Pattern. The LOESS curve (cf., Cleveland, 1979) in Figure 1 (see p. 7) shows the basic pattern of data relating validity to run distance.

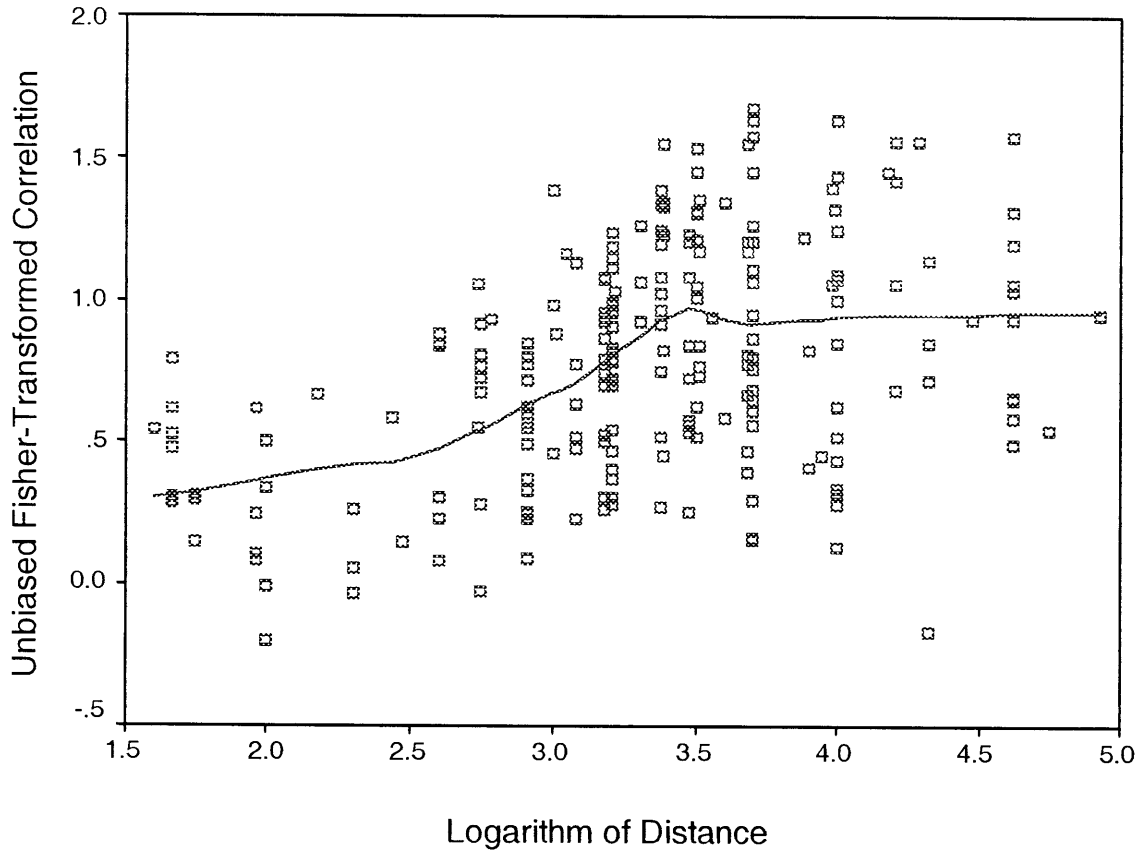


Figure 1
Validity Coefficients as a Function of Distance

Group Classification Models. Fixed-distance tests were grouped several ways to generate group-based models. These models shared the common characteristic that the grouping procedure would explain the observed variation in validity coefficients only if the runs within a group shared a common value. Fixed-distance tests were classified as short- (<1500 m), middle- (1500-1850 m), or long-distance (∃2000 m) runs. One model (S/M/L) treated each category separately. Two other models, (SM/L and S/ML) explored the effect of treating middle-distance runs as short runs or long runs, respectively.

The most extensive group model consisted of 24 groups. Twenty-two (22) groups were specific run distances (e.g., 400 m, 800 m, 5 km). A separate group was included for any run that had been studied in 3 or more samples. The other two groups in this model consisted of all short (<1500 m, n = 9) and long (>1600 m, n = 9) runs that had been studied in only 1 or 2 samples. This model was labeled the "test-by-test" (TxT) model to emphasize

that individual run tests were treated separately when there was enough data to provide a reasonably stable estimate of the $VO2_{2max}$-running performance correlation.

Table 1. Comparison of Group-based Models

Model	df	Π^2	TLI	PTLI
S/ML	1	203.310	.292	.291
SM/L	1	189.051	.271	.270
S/M/L	2	229.561	.325	.322
TxT	23	355.827	.425	.382

Note. S = Short, M = Medium, and L = Long. See text for group definitions. "df" is "degrees of freedom." "TLI" and "PTLI" are the Tucker-Lewis index and the parsimony-adjusted Tucker-Lewis index, respectively. The tabled Π^2s indicate the variation in correlations accounted for by the model. The overall Π^2 was 911.725 with 225 df.

The TxT model clearly was the best group alternative (Table 1). This model was a significant improvement on the next best alternative, the S/M/L model ()Π^2 = 126.27, 21 df, $p < .001$). Even allowing for differences in parsimony, the goodness of fit of the TxT model (PTLI = .382) was better than the S/M/L model fit (PTLI = .322). The S/M/L model was significantly better than either dichotomous model ()Π^2 > 26.25, 1 df, $p < .001$).

Models with Distance as a Continuous Variable. A second set of models used distance as a continuous variable. These models included simple regression and analysis of covariance (ANCOVA) models. The ANCOVA models tested the hypothesis that variations in the size of the correlations within the 2- and 3-group models could be accounted for by distance. If so, it would be unreasonable to treat the tests within a group as equivalent. Preliminary analyses showed that a logarithmic transformation of distance increased the predictive power of the analyses, so this transformation was used in constructing these models.

The analyses led to a mixed model that regressed validity on distance for shorter runs, but treated longer runs as a single group with a common validity (Table 2). A significant amount of variation in the validity coefficients could be accounted for by regressing z_{UF}' on distance (LogDist model; Π^2 > 226.80, PTLI = .324). However, both ANCOVA models improved on this basic regression model ()Π^2 > 17.25, 2 df, $p < .001$). The SM/L model was the better alternative between the two ANCOVA models (SM/L PTLI = .359; S/ML PTLI = .338).

The final mixed model was developed because the regression lines were not parallel for the two SM/L groups (Π^2 = 15.65, 1 df, $p < .001$; cf., Walker & Lev, 1953, pp. 390-393, for the

statistical test). The logarithm of distance predicted z_{UF} in the SM group (Π^2 = 69.724, 1 df, p < .001), but not the L group (Π^2 = 0.08, 1 df, p > .777).

Table 2. Models with Distance as a Continuous Variable.

	Df	Π^2	TLI	PTLI
Log_{10} Distance	1	226.800	.326	.324
S/ML ANCOVA	3	244.05	.342	.338
SM/L ANCOVA	3	258.85	.364	.359
PW	2	258.77	.368	.363

Note. See text for description of models. The within and between values for the PW model indicate the contribution of each model element on total Π^2 for the SM/L ANCOVA.

The mixed model then was constructed based on the ANCOVA results and Figure 1. The model was:

If distance < 2000 m, z_{UF}' = (.225*LogD) - .0036
If distance ∃ 2000 m, z_{UF}' = .9026

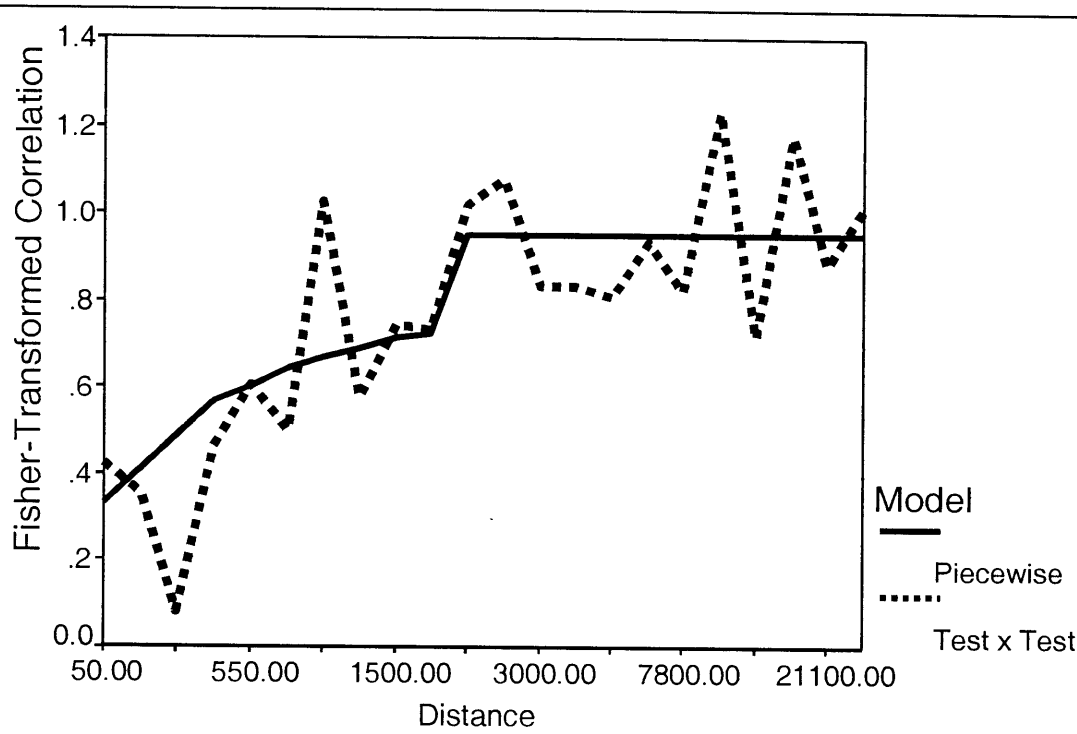
where "LogD" was the logarithm of test distance. This mixed model was labeled the "piecewise" (PW) model because it had distinct prediction components for different ranges of distance. The PW model fit the data almost as well as the full SM/L ANCOVA (Π^2 = 258.85 versus Π^2 = 258.77). The PW PTLI was higher (.359 versus .363).[3]

Comparing the Best Models

The next analysis compared the TxT and PW models as the best alternatives within the two general categories of model. The TxT model fit the data better ()Π^2 > 98.06, 21 df, p < .001), but much of the difference was attributable to the greater parametric complexity of the TxT model. The PTLI values were similar (TxT PTLI = .382; PW PTLI = .363). The sampling variability of PTLI is not known and the specific method of quantifying the parsimony adjustment is only a rule of thumb. Under these conditions, a PTLI difference of .019 was close enough to compare the models further.

Figure 2 compares the model predictions for the 22 run distances that had been studied in 3 or more samples. Differences in the predictions from the two models generally were small. Figure 3 illustrates this fact by expressing the differences as Π^2s. Because the TxT prediction minimizes the weighted squared error for each run distance, Figure 3 also illustrates the loss in predictive accuracy by replacing the TxT model with the PW model.

The effect of a given run test on the overall Π^2 difference between the TxT model and PW models depends on the size of the difference and the sample size for the test (Rosenthal & Rosnow,



Note. Distance is not to scale. Test x Test groups equally spaced.

Figure 2

Piecewise and Test x Test Predictions

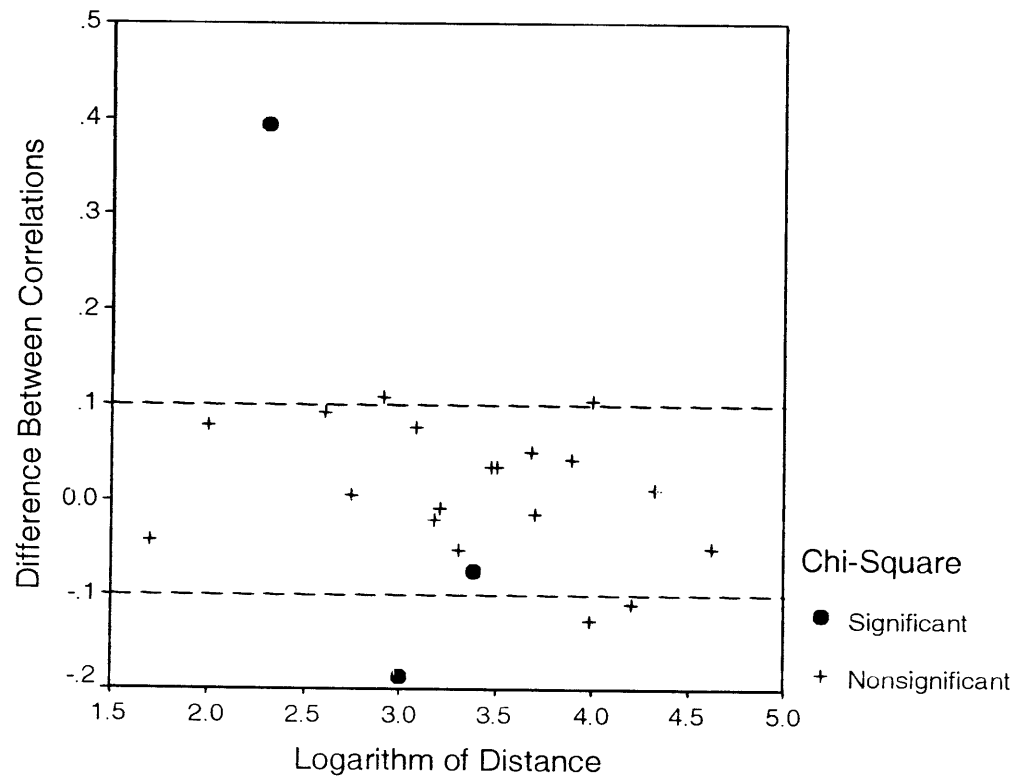


Figure 3

Differences Between Model Predictions

1984). Figure 3 plots the differences between predictions after translating each into a *z*-score, then squaring that score. These computations express each difference as a Π^2 with 1 degree of freedom (Hays, 1963). Stavig and Acock's (1976) procedure was used to determine which Π^2s were greater than expected by chance.

Only 14% (3 of 22) of the differences were greater than expected by chance. One significant difference was of limited practical importance because it was the product of a small effect size combined with a large (N = 650) sample size. Substantial differences between the models were limited to 2 of 22 run distances. The data for these 2 run distances consisted of 7 correlations involving 225 performance scores. For practical purposes, the two models provided effectively equivalent predictions for ~97% of the data (3.4% of 208 correlations; 2.7% of 8,505 performance scores) reviewed. Notice also that both runs for which the showed large significant errors in prediction were ≤ 1 km in length. Thus, neither substantial error was for a run that would be classified as an endurance test in the PW model.

Within-Study Evaluation

The stability of the VO_{2max}-running performance correlation from 2 km on was surprising in light of bioenergetic models. However, the proportion of energy derived from aerobic processes increases relatively slowly for longer runs (Capelli, 1999; di Prampero, et al., 1993; Ward-Smith, 1999). The underlying logic of using bioenergetic models to predict validity trends, therefore, implies that validity will increase slowly for longer runs. If the true validity differences are small, sampling variation and methodological differences between studies could mask the upward trend.

Within-study analyses were conducted to increase the sensitivity of the analyses. Those samples in the data set that performed two or more runs were identified. The VO_{2max}-performance correlations were compared for all pairwise combinations of tests in each sample. Because the people and methods are the same for each correlation in a pair, sampling effects and methods variance are constant. If there is no effect of distance, the comparisons should show that the longer run produced the larger correlation 50% of the time. If the bioenergetic predictions are correct, the longer run should produce the larger correlation more than 50% of the time.

The within-study comparisons were consistent with Figure 1 and the PW model. The correlation for the longer run was larger in 86% (134 of 156) of the pairwise comparisons when at least one run was < 2 km. The longer run produced the larger correlation only 54% (22 of 41) of the time when both runs were ≥ 2 km. The frequency of a larger correlation for the longer run was greater

than chance for short runs (z = 8.97, p < .001) but not long runs (z = 0.47, p > .319).

Random Effects Model. The preceding analyses favored the PW model. Therefore, a random-effects version of that model was computed using Hedges and Vevea's (1998) procedures:

If distance < 2000 m, z_{UF}' = (.259*LogD) - .108
If distance ∃ 2000 m, z_{UF}' = .9518

The random-effects model produced smaller Π^2 values than the fixed-effects model. This trend was expected given the larger variance used to standardize differences. The shift to a random-effects model did not change the inferences about the model components. The regression of z_{UF} on the logarithm of distance still was significant for the shorter runs (Π^2 = 27.423, 1 df, p < .001), but not the longer runs (Π^2 = 0.258, 1 df, p > .611). The difference between the average value for the shorter and longer runs remained significant (Π^2 = 69.776, 1 df, p < .001). The overall model, therefore, was significant (Π^2 = 97.199, 2 df, p < .001)

Fixed-time Tests

A total of 47 fixed-time tests were included in the review. This set included 4 5-min tests, 4 6-min tests, 3 9-min tests, 1 10-min test, 30 12-min tests, and 5 15-min tests. The average validity for the 47 fixed-time tests was r = .752.

Test time, the fixed-time equivalent of test distance for fixed-distance runs, was positively related to correlation magnitude (Π^2 = 61.710, 5 df, p < .001). The average correlations suggested three sets of comparable tests: Set A = {6-min, r = .485}; Set B = {5-min = .659; 9-min, r = .645, 10-min, r = .629}; Set C = {12-min, r = .791; 15-min, r = .835}.

A trend toward higher correlations for longer tests was evident. The trend was most evident as a contrast between tests ∃12 min and tests #10 min. Based on this observation, a model that treated each of the 5 tests as separate groups was compared to two alternatives:

A. Regression: The linear regression of z_{UF} on time was significant (r = .473, Π^2 = 51.586, Y' = .00137*Seconds + .07905).
B. Dichotomous: Short (# 10 min; k = 12) tests were compared to long (∃ 12 min; k = 35) tests. Differences among short tests were nonsignificant (Π^2 = 6.30, 3 df, p > .097). Differences among long tests were nonsignificant (Π^2 = 2.173, 1 df, p > .140). The difference between short and

long tests was highly significant (Π^2 = 53.232, 1 df, p < .001).

The 5-group model predicted better than the time regression model ()Π^2 = 10.124, 4 df, p < .039), but did not improve significantly on the dichotomous model ()Π^2 = 8.478, 4 df, p > .075). Goodness of fit favored the dichotomous model (PTLI = .258) over the regression model (PTLI = .249) and the 5-group model (PTLI = .198).

Boundary Case Analysis for an Endurance Criterion

The PW model and the analyses of fixed-time tests suggested an empirical definition of the term "endurance test." Setting the criterion of ≥2 km or ≥12 minutes provided a reasonable working definition of an endurance test. The definition treats all tests above the distance/time cutoff as equally valid. All tests below the cutoff have lower validity.

The appropriateness of the proposed boundary criteria was evaluated. Two predictions were made for 4 boundary cases, the 1500-meter, 1-mile, 2-kilometer, and 1.5-mile runs. One prediction was based on the validity-distance regression for runs <1500 m (z_{UF}' = .195*Log_{10}D + .0380). The second prediction was the average correlation for tests >1.5 miles (z_{UF}' = .8678). Hoelter's (1983) critical N was used to evaluate the differences (z_{UF} - z_{UF}'). Disch, Frankiewicz and Jackson's (1975) names for their two running performance factors were adopted to label the two predictions "speed" and "distance," respectively.

Table 3. Goodness of Fit for Boundary Tests

		Critical N if Classified as:	
Test	Average z_{UF}	Speed	Distance
1500m	.7442	512	255
1609m	.7317	825	211
2000m	1.0167	38	176
2414m	1.0754	30	93

Note. See text for definition of speed and endurance tests.

The evaluations supported the proposed criteria. The critical N for each proposed classification was 2 to 4 times larger than that for the alternative classification. Larger critical Ns indicate better prediction, so the proposed criteria assigned each boundary test to the portion of the model that provided better predictive accuracy. The critical Ns for the 2-km and 1.5-mi tests were low, but larger for their proposed assignment than for the alternative. In these cases, the fit of the model was not as good as one would like, but the initial classification was the lesser of two evils.

Discussion

This review tested the hypothesis that the validity of run tests as indicators of VO_{2max} increases continuously with distance. The hypothesis was not supported. Validity increased up to 2 km, then remained stable. For fixed-time tests, validity was stable for runs ≥ 12 minutes. The average validity for longer duration runs was comparable to that for longer distance runs. Given this similarity, an endurance run can be defined as any run ≥ 2 km in distance or ≥ 12 minutes in duration.[4] This definition identifies a set of run tests that all possess the same optimal validity as indicators of aerobic capacity.

The recommended definition of endurance runs involves longer runs than some common testing practices (cf., Baumgartner & Jackson, 1982). However, the definition is consistent with Disch et al.'s (1975) factor analysis of performance for run tests ranging from 50 yards to 2 miles. Two factors, "speed" and "distance," were identified. The authors originally classified a 1-mile run as a distance test, but noted that ". . . shorter distance tests of 1 mi or less tended to . . . [load] on both factors, whereas, the distances longer than 1 mi tended to be unidimensional and loaded almost exclusively on the distance run factor" (Disch et al., 1975; p. 169). The shortest distance exceeding 1 mile in their study was 2.01 km (1.25 miles). Thus, proposed criteria are consistent with at least one prior study.

The proposed time criterion for an endurance test is approximate. The 12-minute run is an endurance test. The 9-minute run is not. The optimum criterion might fall between these two values. However, there is too little data on fixed-time runs between 9 and 12 minutes to set the duration criterion with greater precision. Clarification of this issue could be important because time, not distance, is probably the key factor affecting run test validity. For example, Sidney and Shephard's (1977) elderly men and women produced representative validity coefficients for a 12-minute run despite average distances substantially less than 2 km for both groups.

Endurance runs are more valid indicators of aerobic capacity than prior reviews suggest. With the exception of Safrit et al.'s (1988) work, the prior reviews suggest validities in the range of $.60 < r < .65$ (Baumgartner & Jackson, 1982; Katch & Henry, 1972; Knapik, 1989). Safrit et al. (1988) reported a higher value after correcting for measurement error, but their raw correlations were in the range noted in other reviews. In contrast, this review estimates the validity of endurance run tests at $r \approx .74$. The inclusion of shorter runs in the prior reviews is part of the reason for the difference. The analysis of boundary cases showed that even a slight lowering of the criteria

adds run tests with substantially lower correlations, thereby lowering the average.

The proposed endurance criteria do not mean that shorter runs are invalid. The random-effects PW model estimates of validity for shorter runs commonly used to assess aerobic capacity ranged from r = .543 for a 600-yard run to r = .618 for a 1-mile run. Clearly, these tests are not invalid as the correlations are substantially greater than zero. Shorter tests will be useful for estimating aerobic capacity when validities in this range are acceptable and there is some reason to avoid having the study population run the additional distance required to meet the minimum endurance criterion. However, using a shorter run does imply a substantial drop in validity relative to endurance runs. Also, factor loadings from Disch et al.'s (1975) analysis suggest that the estimates of aerobic capacity will be moderately contaminated by differences in anaerobic power.

The endurance criteria are linked to the adoption of the PW model as the best model of the run distance and validity. That decision was supported by the parsimony of the PW model. Adopting the TxT model would increase complexity 1150% (from 2 to 23 parameters) to improve predictions for 9% (2 of 22) of the run tests representing ~3% of the total data. The PW model also has clear connections to current theoretical models of running performance. The increasing validity up to 2 km can be explained by bioenergetics. Critical power, anaerobic threshold, and related physiological concepts (Vandewalle, Vautier, Kachouri, LeChevalier, & Monod, 1997; Walsh, 2000) can account for the range of stable correlations. These concepts predict that there is critical velocity that can be maintained for extended periods of time. Optimal running strategy is to maintain a constant pace that is slightly faster so that anaerobic resources are consumed evenly over the course of the run (Fukuba & Whipp, 1999). Thus, each individual should have an approximately constant pace for longer runs that is determined by aerobic capacity and influenced only slightly by other energetic sources. The implication is that all longer runs are primarily manifestations of a single underlying physiological attribute. From statistical perspective, the tests are congeneric (Lord & Novick, 1968) and should have an approximately constant correlation to the criterion.

The TxT model predictions would be hard to explain physiologically. Mechanisms would have to be identified that could account for an up-and-down pattern of validity coefficients. The pattern might be viewed as a combination of a general upward trend with cyclical variation about that trend that damped to very small fluctuations for longer runs. It is not obvious what physiological constructs could be invoked to explain this pattern.

Noting some limitations of this review puts the results in proper perspective. The conclusions apply with greatest certainty to people between 10 and 50 years of age. Only 4 samples in this review fell outside this range. The risks in generalizing beyond this range may be slight; the 3 samples of older individuals (Sidney & Shephard, 1977; Tanaka, Takeshima, Kato, Niihata, & Ueda, 1990) produced correlations comparable to those for younger people. The statistical significance estimates must be interpreted cautiously. Each correlation was treated as an independent observation even though some were not. More complex computations allowing for the dependencies would yield more precise significance estimates (Becker & Schram, 1994; Steiger, 1978). This limitation is mitigated by the fact that model selection ultimately focused on explanatory precision, not statistical significance. Also, the within-study analysis of correlations provided a qualitative test of the model that allowed for dependencies.

The most important limitation of this review is that validity generalization has not been addressed. Validity was stable for longer runs *on the average*, but there was substantial variation around that average. The variation may indicate that validity is lower for some test populations than for others. Generalizability is critical in the applied use of run tests (Baumgartner & Jackson, 1982; Knapik, 1989; Safrit et al., 1988). This review provides empirical endurance criteria that define a population of run tests that share a common VO_{2max}-running performance correlation. The null hypothesis in generalizability analyses is that different populations of people share a common population correlation. This hypothesis is plausible if run tests are sampled from the population of tests defined by the endurance criteria developed here. The present findings, therefore, provide a starting point for proper selection of correlations suitable for testing generalizability hypotheses. The present findings also identify test type as one factor that affects validity. The average validity of fixed-time endurance tests (r = .797) was significantly higher than that of fixed-distance tests (r = .718, χ^2 = 30.65, 1 df, p < .001). A companion review (Vickers, in preparation) will use the present findings as a point of departure for a detailed exploration of generalizability issues.

The applied uses of the findings can be illustrated by answering the two questions raised in the introduction. "How long does a run have to be to be valid?" If r = .70 were the minimum acceptable validity coefficient, the minimum distance would be 2 km. Reducing the distance by as little as 0.4 km (i.e., to 1 mile) wold incur a significant loss of validity (r = .63). Regarding the second question, "If the current test is 2 km in length, how much will be gained by increasing the distance?", the evidence indicates nothing will be gained. However, they may be some benefit to switching from a fixed-distance test to a fixed-time test. The data also suggest an answer to a third important

applied question, "What is the highest validity that can be achieved with run tests?" The best answer from this review is r = .80. If this validity is not acceptable, some other method of estimating aerobic capacity must be used.

Safrit et al. (1988) concluded that their review of the VO_{2max}-running performance literature provided a framework for future studies. This review has elaborated on the line of study initiated in that paper by developing a quantitative model of the effect of run distance on validity. The model has two important consequences. First, it provides an empirical definition of endurance runs. Second, the model indicates that the validity of run tests as indicators of aerobic capacity is higher than suggested in previous reviews. The empirical definition of an endurance test is a necessary starting point for validity generalization analyses that are the subject of a companion review (Vickers, in preparation). The immediate payoffs from the model developed here include the possibility of making explicit tradeoffs between distance and validity when appropriate. Overall, the PW model should promote better understanding and more effective use of run tests as aerobic capacity indicators.

Footnotes

[1]Safrit et al. (1988) reported two values, r = .771 in the text and r = .741 in Table 1. Whichever value is correct, the analysis procedures corrected for measurement error. The weighted average of the reliability data reported in the paper was r_{xx} = .892 for run tests and r_{yy} = .753 for VO_{2max} measurements. Inserting these values into standard equations to correct for measurement error (Hunter, Schmidt, & Jackson, 1982, p. 54-59), the correction procedure can be reversed to yield the uncorrected correlations:

$$r_{xy} = \Delta_{xy} * SQRT(r_{xx} * r_{yy})$$
$$= .771 * SQRT(.892 * .753)$$
$$= .771 * .820$$
$$= .632$$

or

$$= .608 \text{ (if } \Delta_{xy} = .741).$$

[2]The referent for "validity" has been specified because test validity is the appropriateness of the interpretation of test scores (American Psychological Association, 1985). Most tests have more than one interpretation and, therefore, more than one validity. For example, a run test could be interpreted as performance indicator rather than an estimate of aerobic capacity. This review examines run tests as estimates of aerobic capacity or cardiorespiratory fitness. Unless otherwise indicated, that reference is the sole meaning of validity when the term is used in this paper.

[3]The 2000 meter split point for the group classification may appear too low when examining Figure 1. The graph flattens at a point closer to 2400 meters. This appearance is misleading. LOESS procedures compute the y value for an (x,y) pair by taking a weighted average of observed y values over a range of x values. The weights are larger for data points near the x value than for more distant data points (Cleveland, 1979). The procedure is designed to yield a smoother, robust representation of the data. The resulting graph will be misleading if there are real discontinuities in the data such as that embodied in the PW model. The LOESS approach will yield an artifactually smooth increase in the curve near the transition point. The curve will be smooth and increasing in the transition region because it averages increasing points below the transition with constant points above. As the weights assigned to points in each domain shift, the curve will increase smoothly. The stable value above the transition point will be reached only when the weights assigned to shorter distances all are near zero. This condition will be satisfied only after the x value is well above the actual transition point. Thus, a smooth curve from approximately 2.4 kilometers onward indicates a transition point somewhere below

this value. Other aspects of the analysis indicate that 2 kilometers is reasonable from this perspective.

[4]An apparent conflict between the time and distance definitions of endurance runs should be noted. The average distance covered in a 12-minute run test is 2.5 km, well above the 2.0-km criterion. Regressing average time on average distance for fixed-distance tests, the predicted average time for 2 km is 8:44. This prediction is well below the 12-minute endurance. Note, however, that both predicted criteria refer to average values. The more appropriate reference point might be the time or distance required for the 95th percentile individual. That reference point would be more appropriate given that all individuals have to complete the time or distance in the standard version of the tests. That reference point would be expected to yield closer correspondence between the criteria.

References

*Study contributed 1 or more correlations to the meta-analysis.

*Abe, D., Kazumasa, Y., Yamanobe, K., & Tamura, K. (1998). Assessment of middle-distance running performance in sub-elite young runners using energy cost of running. *European Journal of Applied Physiology, 77*, 320-325.

*Acevedo, E. O., & Goldfarb, A. H. (1989). Increased training intensity effects on plasma lactate, ventilatory threshold, and endurance. *Medicine and Science in Sports and Exercise, 21*(5), 563-568.

American Psychological Association. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago: SmallWaters Corporation.

*Arvey, R. D., Landon, T. E., Nutting, S. M., & Maxwell, S. E. (1992). Development of physical ability tests for police officers: a construct validation approach. *Journal of Applied Psychology, 77*(6), 996-1009.

Baumgartner, T. A., & Jackson, A. S. (1982). *Measurement for evaluation in physical education*. (Second ed.). Dubuque, IA: Wm. C. Brown.

Becker, B. J., & Schram, C. M. (1994). Examining explanatory models through research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 357-382). New York: Russell Sage Foundation.

*Beckett, M. B., & Hodgdon, J. A. (1987). *Lifting and carrying capacities relative to physical fitness measures* (Technical No. 87-26). Naval Health Research Center.

*Bell, A. C., & Hinson, M. M. (1974). Prediction of maximal oxygen intake in women twenty to forty years of age. *Journal of Sports Medicine, 14*, 208-212.

Bentler, P.M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

*Berg, K., & Bell, C. W. (1980). Physiological and anthropometric determinants of mile run time. *Journal of Sports Medicine, 20*, 390-396.

*Billat, V., Renoux, J. C., Pinoteau, J., Petit, B., & Koralsztein, J. P. (1994). Reproducibility of running time to exhaustion at VO_{2max} in subelite runners. *Medicine and Science in Sports and Exercise, 26*(2), 254-257.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

*Brandon, L. J., & Boileau, R. A. (1987). The contribution of selected variables to middle and long distance run performance. *Journal of Sports Medicine, 27*, 157-164.

*Brandon, L. J., & Boileau, R. A. (1992). Influence of metabolic, mechanical and physique variables on middle distance running. *Journal of Sports Medicine and Physical Fitness, 32*(1), 1-9.

*Buono, M. J. (1987). *Validity of the 500 yard swim and 5 kilometer stationary cycle ride as indicators of aerobic fitness* (Technical No. 87-27). Naval Health Research Center.

*Burke, E. J. (1976). Validity of selected laboratory and field tests of physical working capacity. *Research Quarterly, 47*, 95-103.

*Burris, B. (1970). *Reliability and validity of the 12 minute run test for college women*. Paper presented at the AAHPER Convention, Seattle, WA.

Butts, N. K., Henry, B. A., & McLean, D. (1991). Correlations between VO_{2max} and performance times of recreational triathletes. *Journal of Sports Medicine and Physical Fitness, 31*(3), 339-344.

Capelli, C. (1999). Physiological determinants of best performances in human locomotion. *European Journal of Applied Physiology and Occupational Physiology, 80*(4), 298-307.

*Cisar, C. J., Thorland, W. G., Johnson, G. O., & Housh, T. J. (1986). The effect of endurance training on metabolic responses and the prediction of distance running performance. *Journal of Sports Medicine, 26*, 234-240.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*, 829-836.

*Conley, D. S., Cureton, K. J., Hinson, B. T., Higbie, E. J., & Weyand, P. G. (1992). Validation of the 12-minute swim as a field test of peak aerobic power in young women. *Research Quarterly for Exercise and Sport, 63*(2), 153-161.

*Conley, D. L., & Krahenbuhl, G. S. (1980). Running economy and distance running performance of highly trained athletes. *Medicine and Science in Sports and Exercise, 12*(5), 357-360.

*Cooper, K. H. (1968). A means of assessing maximal oxygen intake. *Journal of the American Medical Association, 203*(3), 135-138.

*Costill, D. L. (1967). The relationship between selected physiological variables and distance running performance. *Journal of Sports Medicine and Physical Fitness, 7*, 61-66.

*Costill, D. L., Thomason, H., & Roberts, E. (1973). Fractional utilization of the aerobic capacity during distance running. *Medicine and Science in Sports, 5*(4), 248-252.

*Craig, I. S., & Morgan, D. W. (1998). Relationship between 800-m running performance an accumulated oxygen deficit in middle-distance runners. *Medicine & Science in Sports & Exercise, 30*(11), 1631-1636.

*Cunningham, L. N. (1990). Relationship of running economy, ventilatory threshold, and maximal oxygen consumption to running performance in high school females. *Research Quarterly for Exercise and Sport, 61*(4), 369-374.

*Cureton, K. J., Boileau, R. A., Lohman, T. G., & Misner, J. E. (1977). Determinants of distance running performance in

children: Analysis of a path model. *Research Quarterly, 48*, 270-279.

*Cureton, K. J., Sloniger, M. A., Black, D. M., McCormack, W. P., & Rowe, D. A. (1997). Metabolic determinants of the age-related improvement in one-mile run/walk performance in youth. *Medicine and Science in Sports and Exercise, 29*(2), 259-267.

Cureton, K. J., Sloniger, M. A., O'Bannon, J. P., Black, D. M., & McCormack, W. P. (1995). A generalized equation for prediction of VO2peak from 1-mile run/walk performance. *Medicine and Science in Sports and Exercise, 27*, 445-451.

*Custer, S. J., & Chaloupka, E. C. (1977). Relationship between predicted maximal oxygen consumption and running performance of college females. *Research Quarterly, 48*, 47-50.

*Davies, C. T. M., & Thompson, M. w. (1979). Aerobic performance of female marathon and male ultramarathon athletes. *European Journal of Applied Physiology, 41*, 233-245.

*Deason, J., Powers, S. K., Lawler, J., Ayers, D., & Stuart, M. K. (1991). Physiological correlates to 800 meter running performance. *Journal of Sports Medicine and Physical Fitness, 31*(4), 499-504.

*di Prampero, P. E., Atchou, G., Bruckner, J.-C., & Moia, C. (1986). The energetics of endurance running. *European Journal of Applied Physiology, 55*, 259-266.

di Prampero, P. E., Capelli, C., Pagliaro, P., Antonutto, G., Girardis, M., Samparo, P., & Soule, R. G. (1993). "Energetics of best performances in middle-distance running." *Journal of Applied Physiology, 74*(5), 2318-24.

*Diaz, F. J., Montano, J. G., Melchor, M. T., Guerrero, J. H., & Tovar, J. A. (2000). Validation and reliability of the 1000 meter aerobic test. *Rev Invest Clin, 52*(2), 44-51.

Disch, J., Frankiewicz, R., & Jackson, A. (1975). Construct validation of distance run tests. *Research Quarterly, 46*(2), 169-176

Doolittle, T. L., & Bigbee, R. (1968). The twelve-minute run-walk: a test of cardiorespiratory fitness of adolescent boys. *Research Quarterly, 39*(3), 491-495.

*Drake, V., Jones, G., R., B. J., & Shephard, R. J. (1968). Fitness performance tests and their relationship to the maximal oxygen uptake of adults. *Canadian Medical Association Journal, 99*, 844-848.

*Duggan, A., & Tebbutt, S. D. (1990). Blood lactate at 12 km/h and vOBLA as predictors of run performance in non-endurance athletes. *International Journal of Sports Medicine, 11*(2), 111-115.

Erez, A., Bloom, M. C., & Wells, M. T. (1996). Using random rather than fixed effects models in meta-analysis: implications for situational specificity and validity generalization. *Personnel Psychology, 49*, 275-306.

*Evans, S. L., Davy, K. P., Stevenson, E. T., & Seals, D. R. (1995). Physiological determinants of 10-km performance in

highly trained female runners of different ages. *Journal of Applied Physiology, 78*(5), 1931-1941.

*Falls, H. B., Ismail, A. H., & MacLeod, D. F. (1966). Estimation of Maximum oxygen uptake in adults from AAHPER youth fitness test items. *Research Quarterly, 37*(2), 192-201.

*Farrell, P. A., Wilmore, J. H., Coyle, E. F., Billing, J. E., & Costill, D. L. (1979). Plasma lactate accumulation and distance running performance. *Medicine and Science in Sports, 11*(4), 338-344.

*Fay, L., Londeree, B. R., LaFontaine, T. P., & Volek, M. R. (1989). Physiological parameters related to distance running performance in female athletes. *Medicine and Science in Sports and Exercise, 21*(3), 319-324.

*Florence, S.-L., & Weir, J. P. (1997). Relationship of critical velocity to marathon running performance. *European Journal of Applied Physiology, 75*, 274-278.

*Foster, C., Costill, D. L., Daniels, J. T., & Fink, W. J. (1978). Skeletal muscle enzyme activity, fiber composition, and VO2 max in relation to distance running performance. *European Journal of Applied Physiology, 39*, 73-80.

*Forster, C., Daniels, J. T., & Yarbrough, R. A. (1977). Physiological and training correlates of marathon running performance. *Australian Journal of Sports Medicine, 9*, 58-61.

Fukuba, Y., & Whipp, B. J. (1999). A metabolic limit on the ability to make up for lost time in endurance events. *Journal of Applied Physiology, 87*, 853-861.

*Getchell, L. H., Kirkendall, D., & Robbins, G. (1977). Prediction of maximal oxygen uptake in young adult women joggers. *Research Quarterly, 48*, 61-67.

*Grant, J. A., Joseph, A. N., & Campagna, P. D. (1999). The prediction of VO_{2max}: a comparison of 7 indirect tests of aerobic power. *Journal of Strength and Conditioning Research, 13*(4), 346-352.

*Grant, S., Corbett, K., Amjad, A. M., Wilson, J., & Aitchison, T. (1995). A comparison of methods of predicting maximum oxygen uptake. *British Journal of Sports Medicine, 29*, 147-152.

*Gutin, B., Fogle, R. K., & Stewart, K. (1976). Relationship among submaximal heart rate, aerobic power, and running performance in children. *Research Quarterly, 47*, 536-540.

*Gutin, B., Trinidad, A., Norton, C., Giles, E., Giles, A., & Stewart, K. (1978). Morphological and physiological factors related to endurance performance of 11- to 12-year-old girls. *Research Quarterly, 49*, 44-52.

*Hagan, R. D., Smith, M. G., & Gettman, L. R. (1981). Marathon performance in relation to maximal aerobic power and training indices. *Medicine and Science in Sports and Exercise, 13*(3), 185-189.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.

*Harrison, M. H., Bruce, D. L., Brown, G. A., & Cochrane, L. A. (1980). A comparison of some indirect methods for predicting maximal oxygen uptake. *Aviation, Space, and Environmental Medicine, 51*(10), 1128-1133.

*Hazard, A. A. (1982). *The effects of endurance training at 2,440m altitude on maximal oxygen uptake at altitude and sea level in young male and female middle distance runners.* Master's, San Diego State University.

Hays, W. L. (1963). *Statistics for Psychologists.* New York: Holt, Rinehart, Winston.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486-504.

*Hodgdon, J. A., Vickers, R. R., Jr., & Bennett, B. L. (1983). *Impact of physiological and psychological factors on performance in a middle-distance run* (Technical Report No. 80-30). Naval Health Research Center.

Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods and Research, 11*, 325-344.

*Houmard, J. A., Craig, M. W., O'Brien, K. F., Smith, L. L., Israel, R. G., & Wheeler, W. S. (1991). Peak running velocity, submaximal energy expenditure, VO_{2max}, and 8 km distance running performance. *Journal of Sports Medicine and Physical Fitness, 31*(3), 345-350.

*Huhn, R. R. (1975). *The reliability, validity, and predictability of twelve- and fifteen-minute field tests in relation to laboratory maximal oxygen uptake tests.* Masters, San Diego State University.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis.* Newbury Park: Sage.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis.* Beverly Hills, CA: Sage.

*Iwaoka, K., Hatta, H., Atomi, Y., & Miyashita, M. (1988). Lactate, respiratory compensation thresholds, and distance running performance in runners of both sexes. *International Journal of Sports Medicine, 9*(5), 306-309.

*Jackson, A., der Weduwe, K., Schick, R., & Sanchez, R. (1990). An analysis of the validity of the three-mile run as a field test of aerobic capacity in college males. *Research Quarterly, 61*(3), 233-237.

*Jackson, A. S., & Coleman, A. E. (1976). Validation of distance run tests for elementary school children. *Research Quarterly, 47*, 86-94.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: assumptions, models, and data.* Beverly Hills, CA: Sage.

*Johnson, D. J., Oliver, R. A., & Terry, J. W. (1979). Regression equation for prediction of performance in the twelve minute run walk test. *Journal of Sports Medicine, 19*, 165-170.

*Katch, V., & Henry, F. M. (1972). Prediction of running performance from maximal oxygen debt and intake. *Medicine and Science in Sports, 4*(4), 1887-191.
*Katch, F. I., McArdle, W. D., Czula, R., & Pechar, G. S. (1973). Maximal oxygen intake, endurance running performance, and body composition in college women. *Research Quarterly, 44*(3), 301-312.
*Katch, F. I., Pechar, G. S., McArdle, W. D., & Weltman, A. L. (1973). Relationship between individual differences in a steady pace endurance running performance and maximal oxygen intake. *Research Quarterly, 44*(2), 206-215.
*Katch, V.I. (1970). The role of maximal oxygen intake in endurance performance. Paper presented at the AAHPER Convention, Seattle, WA. (Cited in Safrit et al., 1988).
*Kearney, J. T., & Byrnes, W. C. (1974). Relationship between running performance and predicted maximum oxygen uptake among divergent ability groups. *Research Quarterly, 45*(1), 9-15.
*Kitagawa, K., Miyashita, M., & Yamamoto, K. (1977). Maximal oxygen uptake, body composition, and running performance in young Japanese adults of both sexes. *Japanese Journal of Physical Education, 21*(6), 335-340.
Knapik, J. (1989). The Army Physical Fitness Tests (APFT): a review of the literature. *Military Medicine, 154*, 326-329.
Kohrt, W. M., Morgan, D. W., Bates, B., & Skinner, J. S. (1987). Physiological responses of triathletes to maximal swimming, cycling, and running. *Medicine and Science in Sports and Exercise, 19*(1), 51-55.
Kohrt, W. M., O'Connor, J. S., & Skinner, J. S. (1989). Longitudinal assessment of responses by triathletes to swimming, cycling, and running. *Medicine and Science in Sports and Exercise, 21*(5), 569-575.
*Krahenbuhl, G. S., Pangrazi, R. P., Burkett, L. N., Schneider, M. J., & Petersen, G. (1977). Field estimation of VO_{2max} in children eight years of age. *Medicine and Science in Sports, 9*(1), 37-40.
*Krahenbuhl, G. S., Pangrazi, R. P., Petersen, G. W., Burkett, L. N., & Schneider, M. J. (1978). Field testing of cardiorespiratory fitness in primary school children. *Medicine and Science in Sports, 10*(3), 208-213.
Krahenbuhl, G. S., Wells, C. L., Brown, C. H., & Ward, P. E. (1979). Characteristics of national and world class female pentathletes. *Medicine and Science in Sports, 11*(1), 20-23.
*Kumagai, S., Tanaka, K., Matsuura, Y., Matsuzaka, A., Hiraboba, K., & Asano, K. (1982). Relationships of the anaerobic threshold with the 5 km, 10 km, and 10 mile races. *European Journal of Applied Physiology, 49*, 13-23.
*Lacour, J. R., Padilla-Magunacelaya, S., Barthelemy, J. C., & Dormois, D. (1990). The energetics of middle-distance running. *European Journal of Applied Physiology, 60*, 38-43.
*Lacour, J. R., Padilla-Magunacelaya, S., Chatard, J. C., Arsac, L., & Barthelemy, J. C. (1991). Assessment of running

velocity at maximal oxygen uptake. *European Journal of Applied Psychology, 62*, 77-82.

*Leach, D. A. (1983). *The measurement of cardio-respiratory endurance and the Standard Evaluation for Army personnel in the 40-45 age category*. Carlisle, PA: U.S. Army War College, Study Project. (Cited in Knapik, 1989).

*Lehmann, M., Berg, A., Kapp, R., Wessinhage, T., & Keul, J. (1983). Correlations between laboratory testing and distance running performance in marathoners of similar performance ability. *International Journal of Sports Medicine, 4*(4), 226-230.

*Loftin, M., Zingraf, S., Warren, B., Jones, C. J., Brandon, J. E., & Harsha, D. (1986). Influence of physiological function and perceptual effort on 1.5 miles performance in college women. *Journal of Sports Medicine, 26*, 214-218.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.

*MacNaughton, L., Croft, R., Pennicott, J., & Long, T. (1990). The 5 and 15 minute runs as predictors of aerobic capacity in high school students. *Journal of Sports Medicine and Physical Fitness, 30*, 24-28.

*Mahon, A. D., Del Corral, P., How, C. A., Duncan, G. E., & Ray, M. L. (1996). Physiological correlates of 3-kilometer running performance in male children. *International Journal of Sports Medicine, 17*(8), 580-584.

*Maksud, M. G., Cannistra, C., & Dublinski, D. (1976). Energy expenditure and VO_{2max} of female athletes during treadmill exercise. *Research Quarterly, 47*, 692-697.

*Maksud, M. G., & Coutts, K. D. (1971). Application of the Cooper twelve-minute run-walk test to young males. *Research Quarterly, 42*(1), 54-59.

*Massicotte, D. R., Gauthier, R., & Markon, P. (1985). Prediction of VO_{2max} from the running performance in children aged 10-17 years. *Journal of Sports Medicine, 25*, 10-17.

*Matsui, H., Miyashita, M., Miura, M., Kobayoshi, K., Hoshikawa, T., & Kamei, S. (1972). Maximum oxygen intake and its relationship to body weight of Japanese adolescents. *Medicine and Science in Sports, 4*(1), 29-32.

*Mayhew, J. L., & Andrew, J. (1975). Assessment of running performance in college males from aerobic capacity percentage utilization coefficients. *Journal of Sports Medicine, 15*, 342-345.

McArdle, W.D., Katch, F. I., & Katch, V. L. (1996). *Exercise Physiology* (Fourth Ed.). Baltimore: Williams & Wilkins.

*McCormack, W. P., Cureton, K. J., Bullock, T. A., & Weyand, p. G. (1991). Metabolic determinants of 1-mile run/walk performance in children. *Medicine and Science in Sports and Exercise, 23*(5), 611-617.

*McCutcheon, M. C., Stichar, S. A., Giese, M. D., & Nagle, F. J. (1990). A further analysis of the 12-minute run prediction of maximal aerobic power. *Research Quarterly for Exercise and Sport, 61*(3), 280-283.

*McNaughton, L., Hall, P., & Cooley, D. (1998). Validation of several methods of estimating maximal oxygen uptake in young men. *Perceptual and Motor Skills, 87,* 575-584.
*Mello, R. P., Murphy, M. M., & Vogel, J. A. (1984). *Relationship between the Army two mile run test and maximal oxygen uptake* (Technical No. T3/85). U.S. Army Research Institute of Environmental Medicine.
*Metz, K. F., & Alexander, J. F. (1970). An investigation of the relationship between maximum aerobic work capacity and physical fitness in twelve- to fifteen-year-old boys. *Research Quarterly, 41*(1), 75-81.
*Morgan, D. W., Baldini, F. D., Martin, P. E., & Kohrt, W. M. (1989). Ten kilometer performance and predicted velocity at VO_{2max} among well-trained male runners. *Medicine and Science in Sports and Exercise, 21*(1), 78-83.
Morrison, D., & Henkel, R. E. (1970). *The great significance test controversy*. Chicago: Aldine.
Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105*(3), 430-445.
*Myles, W. S., Brown, T. E., & Pope, J. I. (1980). A reassessment of a running test as a measure of cardiorespiratory fitness. *Ergonomics, 23*(6), 543-547.
*Noakes, T. D., Myburgh, K. H., & Schall, R. (1990). Peak treadmill running velocity during the VO_{2max} test predicts running performance. *Journal of Sports Sciences, 8*(35-45).
*O'Donnell, C., Smith, D. A., O'Donnell, T. V., & Stacy, R. J. (1984). Physical fitness of New Zealand army personnel; correlation between field tests and direct laboratory assessments--anaerobic threshold and maximum O_2 uptake. *New Zealand Medical Journal, 97*(760), 476-479.
*O'Gorman, D., Hunter, A., McDonnacha, C., & Kirwan, J. P. (2000). Validity of field tests for evaluating endurance capacity in competitive and international-level sports participants. *Journal of Strength and Conditioning Research, 14*(1), 62-67.
Olkin, I., & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Annals of Mathematical Statistics, 29,* 201-211.
*Paavolinen, L. M., Nummela, A. T., & Rusko, H. K. (1999). Neuromuscular characteristics and muscle power as determinants of 5-km running performance. *Medicine & Science in Sports & Exercise, 31*(1), 124-130.
*Padilla, S., Bourdin, M., Barthelemy, J. C., & Lacour, J. R. (1992). Physiological correlates of middle-distance running performance. *European Journal of Applied Physiology and Occupational* Physiology, *65,* 561-566.
*Palgi, Y., & Gutin, B. (1984). Running performance in children. *Medicine and Science in Sports and Exercise, 16,* 158.
*Palgi, Y., Gutin, B., Young, J., & Alehandro, D. (1984). Physiologic and anthropometric factors underlying endurance

performance in children. *International Journal of Sports Medicine, 5*(2), 67-73.
*Peronnet, F., Thibault, G., Rhodes, E. C., & McKenzie, D. C. (1987). Correlation between ventilatory threshold and endurance capability in marathon runners. *Medicine and Science in Sports and Exercise, 19*(6), 610-615.
Popper, K. R. (1959). *The logic of scientific discovery.* N.Y.: Basic Books.
*Powers, S. K., Dodd, S., Deason, R., Byrd, R., & McKnight, T. (1983). Ventilatory threshold, running economy and distance running performance of trained athletes. *Research Quarterly for Exercise and Sport, 54*(2), 179-182.
*Ramsbottom, R., Nute, M. G. L., & Williams, C. (1987). Determinants of five kilometre running performance in active men and women. *British Journal of Sports Medicine, 21*(2), 9-13.
*Ramsbottom, R., Williams, C., Boobis, L., & Freeman, W. (1989). Aerobic fitness and running performance of male and female recreational runners. *Journal of Sports Sciences, 7*, 9-20.
*Ramsbottom, R., Williams, C., Fleming, N., & Nute, M. L. G. (1989). Training induced physiological and metabolic changes associated with improvements in running performance. *British Journal of Sports Medicine, 23*(3), 171-176.
*Rasch, P. J. (1974). Maximal oxygen intake as a predictor of performance in running events. *Journal of Sports Medicine, 14*, 32-39.
*Rasch, P. J., & Wilson, D. (1964). The correlation of selected laboratory tests of physical fitness with military endurance. *Military Medicine*, 256-258.
Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 301-322). New York: Russell Sage Foundation.
*Ribisl, P. M., & Kachadorian, W. A. (1969). Maximal oxygen intake prediction in young and middle-aged males. *Journal of Sports Medicine, 9*, 17-22.
Rosenthal, R. (1984). Meta-analytic procedures for social research. Beverly Hills, CA: Sage.
Rosenthal, R., & DiMatteo, R. (2001). Meta-analysis: recent developments in quantitative methods for literature reviews. In S. T. Fiske, D. L. Schacter, & C. Zahn-Waxler (eds.), *Annual review of psychology*, Vol. 52 (pp. 59-82). Palo Alto, CA: Annual Reviews, Inc.
Rosenthal, R., & Rosnow, R. L. (1984). *Essentials of behavioral research.* New York: McGraw-Hill.
*Rotstein, A., Dotan, R., Bar-Or, O., & Tenenbaum, G. (1986). Effect of training on anaerobic threshold, maximal aerobic power and anaerobic performance of preadolescent boys. *International Journal of Sports Medicine*, 7(5), 281-286.
*Rudzki, S. J. (1989). Weight-load marching as a method of conditioning Australian army recruits. *Military Medicine, 154*(4), 201-205.

Safrit, M. J., Hooper, L. M., Ehlert, S. A., Costa, M. G., & Patterson, P. (1988). The validity generalization of distance run tests. *Canadian Journal of Sports Science, 13*(4), 188-196.

*Salzer, D. W. (1996). *The relationship of strength to endurance while exercising in a chemical warfare uniform in the heat.* Masters, San Diego State University.

Schabort, E. J., Killian, S. C., St Clair Gibson, A., Hawley, J. A. & Noakes, T. D. (2000). Prediction of triathlon race time from laboratory testing in national triathletes. *Medicine and Science in Sports and Exercise, 32,* 844-849.

*Scrimgeour, A. G., Noakes, T. D., Adams, B., & Myburgh, K. (1986). The influence of weekly training distance on fractional utilization of maximum aerobic capacity in marathon and ultramarathon runners. *European Journal of Applied Physiology, 55,* 202-209.

*Shaver, L. G. (1975). Maximum aerobic power and anaerobic work capacity prediction from various running performances of untrained college men. *Journal of Sports Medicine, 15,* 147-150.

*Sidney, K. H., & Shephard, R. J. (1977). Maximum and submaximum exercise tests in men and women in the seventh, eighth, and ninth decades of life. *Journal of Applied Physiology: Respiratory, Environmental, and Exercise Physiology, 43*(2), 280-287.

*Sjodin, B., & Svedenhag, J. (1985). Applied physiology of marathon running. *Sports Medicine, 2,* 83-99.

*Sloniger, M. A., Cureton, K. J., & O-Bannon, P. J. (1997). One-mile run-walk performance in young men and women: role of anaerobic metabolism. *Canadian Journal of Applied Physiology, 22*(3), 337-350.

*Sparling, P. B., & Cureton, K. J. (1983). Biological determinants of the sex difference in 12-min run performance. *Medicine and Science in Sports and Exercise, 15*(3), 218-223.

Spencer, M. R., & Gastin, P.B. (2001). Energy system contribution during 200- to 1500-m running in highly trained athletes. *Medicine and Science in Sports and Exercise, 33,* 157-162.

SPSS, Inc. (1998a). *SPSS Base 8.0 Applications Guide.* Chicago: SPSS, Inc.

SPSS, Inc. (1998b). *SPSS Advanced Statistics.* Chicago: SPSS, Inc.

Stavig, G. R., & Acock, A. C. (1976). Evaluating the degree of dependence for a set of correlations. *Psychological Bulletin, 83,* 236-241.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245-251.

*Tanaka, K., & Matsuura, Y. (1982). A multivariate analysis of the role of certain anthropometric and physiological attributes in distance running. *Annals of Human Biology, 9*(5), 473-482.

*Tanaka, K., Matsuura, Y., Kumagai, S., Matsuzaka, A., Hirakoba, K., & Asano, K. (1983). Relationships of anaerobic

threshold and onset of blood lactate accumulation with endurance performance. *European Journal of Applied Physiology, 52*, 51-56.

*Tanaka, K., Takeshima, N., Kato, T., Niihata, S., & Ueda, K. (1990). Critical determinants of endurance performance in middle-aged and elderly endurance runners with heterogeneous training habits. *European Journal of Applied Physiology, 59*, 443-449.

*Tanaka, K., Watanabe, H., Konishi, t., Mitsuzono, R., Sumida, S., Tanaka, S., Fukuda, T., & Nakadomo, F. (1986). Longitudinal associations between anaerobic threshold and distance running performance. *European Journal of Applied Physiology, 55*, 248-252.

*Trone, D. W. (1989). *Predicting 10km run performance time from physiological measurements*. Masters, San Diego State University.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10.

*Unnithan, V. B., Timmons, J. A., Paton, J. Y., & Rowland, T. W. (1995). Physiological correlates to running performance in pre-pubertal distance runners. *International Journal of Sports Medicine, 16*(8), 528-533.

Vandewalle, H., Vautier, J. F., Kachouri, M., LeChevalier, J.-M., & Monod, H. (1997). Work-exhaustion time relationship and the critical power concept: a critical review. *Journal of Sports Medicine & Physical Fitness, 37*, 89-102.

*van Mechelen, W., Hlobil, H., & Kemper, H. C. G. (1986). Validation of two running tests as estimates of maximal aerobic power in children. *European Journal of Applied Physiology, 55*, 503-506.

*Vodak, P. A., & Wilmore, J. H. (1975). Validity of the 6-minute jog-walk and the 600-yard run-walk in estimating endurance capacity in boys, 9-12 years of age. *Research Quarterly for Exercise and Sport, 46*(2), 230-234.

*Wannamaker, G. S. (1970). A study of the validity and reliability of 12-minute run under selected motivational conditions. *American Corrective Therapy Journal, 24*(3), 69-72.

Walker, H. M., & Lev, J. (1953). *Statistical inference*. New York: Holt, Rinehart, and Winston.

Walsh, M. L. (2000). Whole body fatigue and critical power: a physiological interpretation. *Sports Medicine, 29*, 153-166.

Wanous, J. P., Sullivan, S. E., & Malinak, J. (1989). The role of judmgnet calls in meta-analysis. *Journal of Applied Psychology, 74*(2), 259-264.

Ward-Smith, A. J. (1999). Aerobic and anaerobic energy conversion during high-intensity exercise. *Medicine and Science in Sports and Exercise, 31*, 1855-1860.

*Weyand, P. G., Cureton, K. J., Conley, D. S., Sloniger, M. A., & Liu, Y. L. (1994). Peak oxygen deficit predicts sprint and

middle-distance track performance. *Medicine and Science in Sports and Exercise, 26*(9), 1174-1180.

White, H. D. (1994). Scientific communication and literature retrieval. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 41-56). New York: Russell Sage Foundation.

*Wiley, J. F., & Shaver, L. G. (1972). Prediction of maximum oxygen intake from running performance of untrained young men. *Research Quarterly, 43*(1), 89-93.

*Wyndham, C. H., Strydom, N. B., van Graan, C. H., van Rensburg, A. J., Rogers, G. G., Greyson, J. S., & van der Walt, W. H. (1971). Walk or jog for health: II. Estimating the maximum aerobic capacity for exercise. *South African Medical Journal, 45*, 53-57.

*Yoshida, T., Chida, M., Ichioka, M., & Suda, Y. (1987). Blood lactate parameters related to aerobic capacity and endurance performance. *European Journal of Applied Physiology, 56*, 7-11.

*Yoshida, T., Ishiko, T., & Muraoka, I. (1983). Cariorespiratory functions in children with high and low performances in endurance running. *European Journal of Applied Physiology, 51*, 313-319.

*Yoshida, T., Udo, M., Iwai, K., Chida, M., Ichioka, M., Nakadomo, F., & Yamaguchi, T. (1990). Significance of the contribution of aerobic and anaerobic components to several distance running performances in female athletes. *European Journal of Applied Physiology, 60*, 249-253.

*Zacharogiannis, E., & Farrally, M. (1993). Ventilatory threshold, heart rate deflection point and middle distance running performance. *Journal of Sports Medicine and Physical Fitness, 33*(4), 337-347.

Zhou, S., Robson, S. J., King, M. J., & Davie, A. J. (1997). Correlations between short-course triathlon performance and physiological variables determined in laboratory cycle and treadmill tests. *Journal of sports Medicine and Physical Fitness, 37*(2), 122-130.

*Zwiren, L. D., Freedson, P. S., Ward, A., Wilke, S., & Rippe, J. M. (1991). Estimation of VO_{2max}: a comparative analysis of five exercise tests. *Research Quarterly for Exercise and Sport, 62*(1), 73-78.

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE & DATE COVERED
	1 Jul 01	Technical Report 01/01/00 – 07/01/01

4. TITLE AND SUBTITLE
Running Performance as an Indicator of VO_{2max}: Distance Effects

5. FUNDING NUMBERS
Program Element: 63706N
Work Unit Number: M0096.001-6417
USAMMRC Reimbursable 60109

6. AUTHOR(S)
Vickers, Ross R., Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

8. PERFORMING ORGANIZATION
.
Report

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
Office of Naval Research
800 North Quincy St.
Arlington, VA 22217-5600

Chief, Bureau of Medicine and Surgery
Code: BUMED-26
2300 E Street NW
Washington, DC 20372-5300

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT *(Maximum 200 words)*

Distance runs often are used to estimate aerobic capacity. This meta-analysis of 226 correlations from 122 studies involving for fixed-distance run tests produced a quantitative model of run test validity as a function of distance. Validity, the correlation between maximum oxygen uptake (VO_{2max}) and running performance, increased with logarithm of distance up to 2 km. Validity was stable at $r = .718$ for runs ≥ 2 km. Based on these results, 2 km is an empirical minimum distance criterion for classifying a run as an endurance test. Analysis of a smaller set of 47 correlations for fixed-time run tests indicated that runs ≥ 12 minutes had a similar correlation ($r = .797$) and should be considered endurance runs. Runs that meet or exceed these minimum distance and time criteria provide interchangeable estimates of aerobic capacity.

14. SUBJECT TERMS
run tests aerobic capacity validity modeling

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unclassified

NSN 7540-01-280-550

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102